

Approximation Methods for Conceptual Design of Complex Systems

Wu Li and Sharon Padula
NASA Langley Research Center



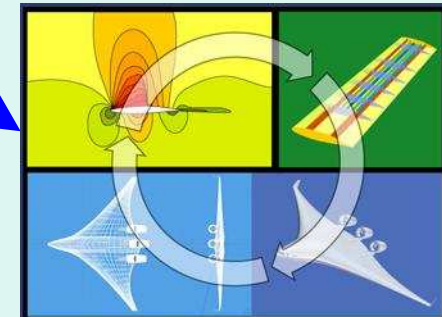
Eleventh International Conference on Approximation Theory
Gatlinburg, Tennessee, May 18-22, 2004

**Approximation of Functions
(Numerical Analysis)**

**Regression Methods
(Statistics)**

**Data Fitting
(Engineering)**

**Systems Analysts &
Method Developers**



**Conceptual Design of
Complex Systems**

Outline

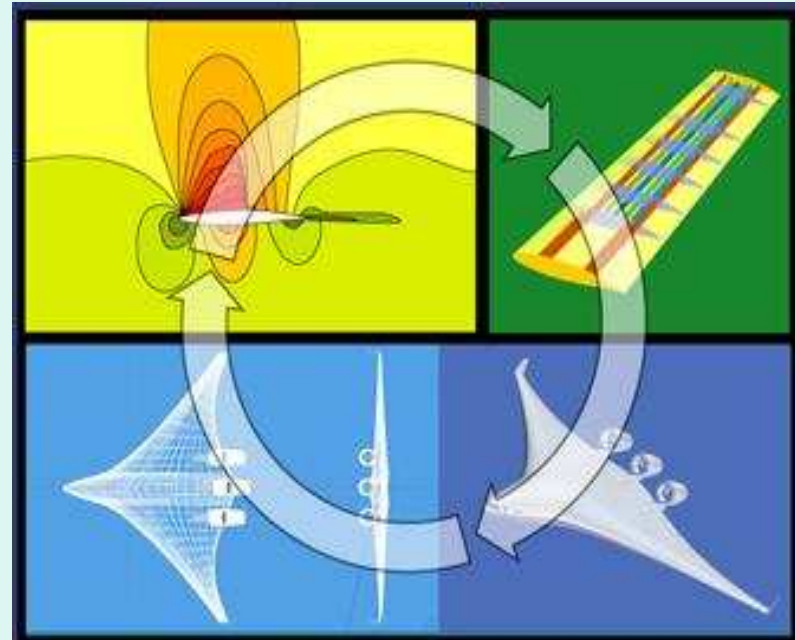
- Needs for approximation in conceptual design of complex systems
- Curse of dimensionality and variable screening
- Overview of approximation methods
 - Interpolation: Kriging and radial basis functions
 - Local regression for smooth approximation
 - Robust data fitting
 - Ridge regression to avoid overfitting
 - Fitting data in feature space for irregularly distributed data
 - Variable-fidelity approximation for calibration of low-fidelity models using high-fidelity data
- Selection of model parameters
- Visualization in high-dimensional space
- Conclusions and future research problems

Needs for Approximation in Conceptual Design

- **Historical or design data fitting**
- **Calibration of low-fidelity model**
- **Capture second order trend with a few data points**
- **Global smooth approximation**

Aircraft Conceptual Design

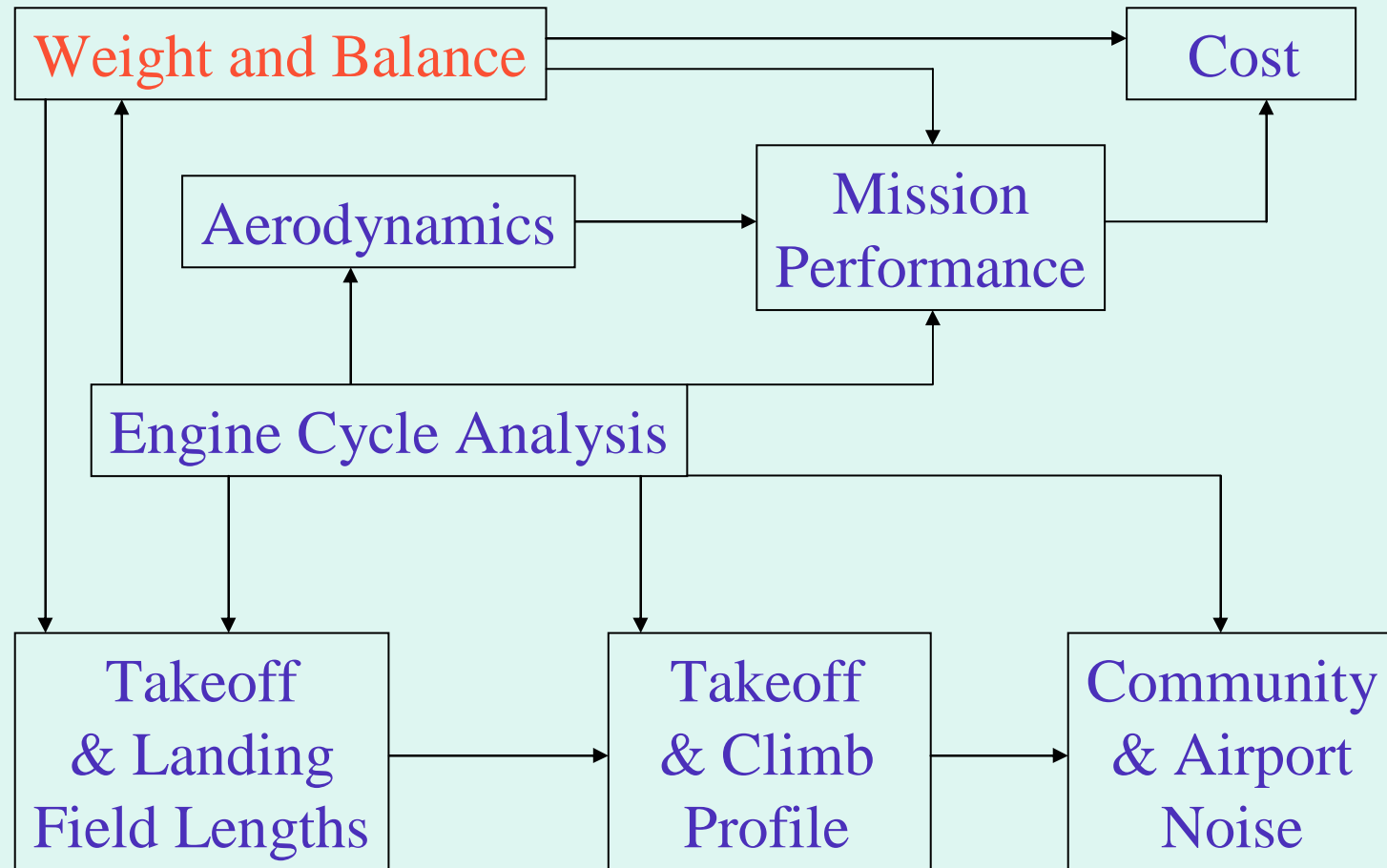
- Will the concept work?
- What does it look like?
- What requirements drive the design?
- What trade-offs should be considered?
- What should it weigh and cost?



Daniel P. Raymer, "Aircraft Conceptual Design"
2nd Edition, AIAA Education Series (1992).

**NASA Applications: Technology assessment
and revolutionary vehicle concept evaluation**

Conceptual Design Requires Multidisciplinary Analysis

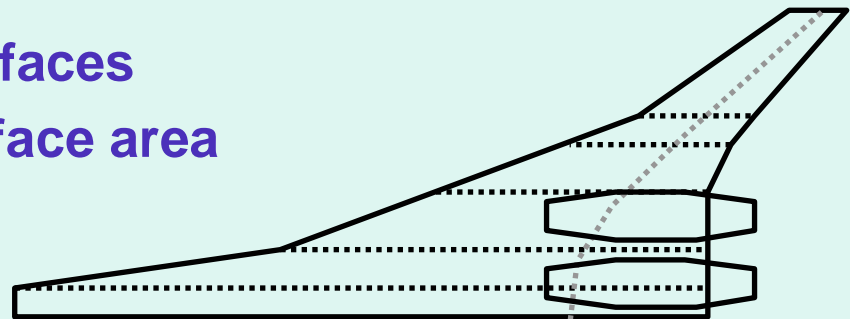


Time constraint: weeks to months for conceptual design

Need for Approximation Methods

Wing Weight Estimation Example (by McCullers)

- Bending material
 - Low-fidelity structural analysis for stress sizing
 - Approximate aerodynamic forces on wing
 - Approximate structure with plates: variable thickness, chord, and sweep angle
 - Approximate engines with masses
 - Response surfaces for constraints like flutter and divergence
 - Correlation with existing aircraft for ideal versus “as-built” weights
- Spars, ribs, and control surfaces
 - Depends on control surface area
- Miscellaneous wing weight
 - Depends on wing area



Two Basic Needs for Approximation

RSM is a popular choice for approximation in engineering simulation. But RSM works best with design of experiments for a local approximation and can not meet the critical needs for approximation in conceptual design.

- Need accurate approximations of historical data that are scarce and poorly distributed in a high-dimensional space.
- Need calibration of a low-fidelity simulation code to match high-fidelity data.

Outline

- Needs for approximation in conceptual design of complex systems
- Curse of dimensionality and variable screening
- Overview of approximation methods
 - Interpolation: Kriging and radial basis functions
 - Local regression for smooth approximation
 - Robust data fitting
 - Ridge regression to avoid overfitting
 - Fitting data in feature space for irregularly distributed data
 - Variable-fidelity approximation for calibration of low-fidelity models using high-fidelity data
- Selection of model parameters
- Visualization in high-dimensional space
- Conclusions and future research problems

Curse of Dimensionality

- The curse comes from the fact that given a finite number of data points, the high-dimensional unit cube is mostly empty. (Friedman and Stuetzle, 1981)
- The curse implies that unless you have an enormous number of data points generated by a simulation model, any data fitting method will not get a good approximation of the true response of the simulation model.

Bottom Line: Forget about good fitting in high-dimensional spaces.

Variable Screening

Traditional approach

1. Evaluate the response with a 2-level fractional factorial design.
2. Identify main effects that account for most of the variance in the response.

Adaptive variable screening

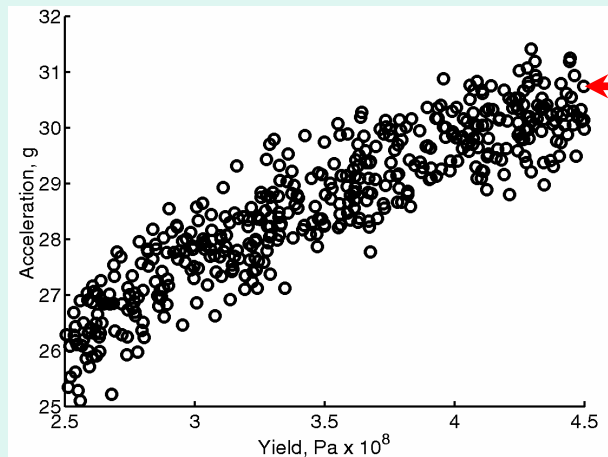
1. Randomly divide variables into groups.
2. Determine main effects and quadratic effects between the groups.
3. Eliminate any group of variables with small effect.
4. Repeat from step 1 with remaining variables until analyst is satisfied.

Adaptive Variable Screening Technique

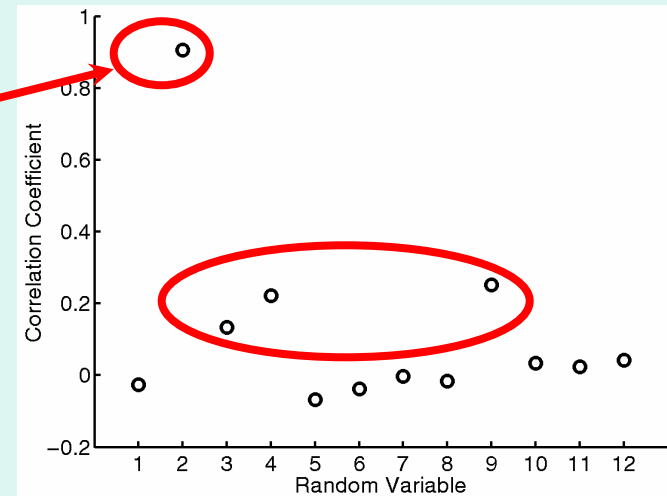
- Typical results: If there are a large number k of variables but only a few important ones, then group screening requires $m \ll k$ analyses.
- Assumptions:
 - Fewer analyses are needed to rank by importance than to quantify importance.
 - A knowledgeable user can identify the true important variables from a ranked list of variables, even if some insignificant variables are misclassified as important ones.

Variable Screening (Lyle, Stockwell, & Hardy, 2003)

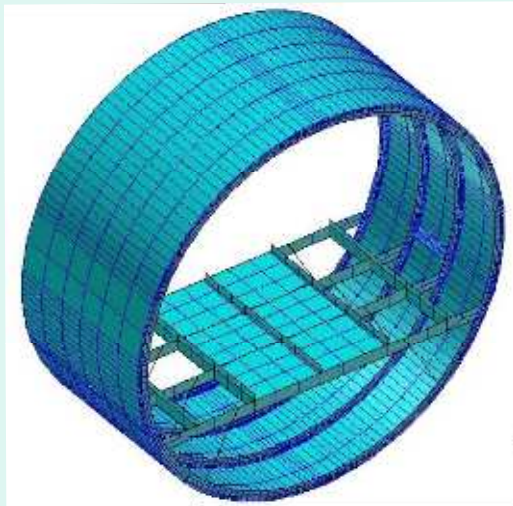
Four variables are highly correlated with crash survivability



1. DOE with low-fidelity analysis



2. Identify four important variables



3. High-fidelity simulation with reduced number of variables



4. Drop test instrumented fuselage

Outline

- Needs for approximation in conceptual design of complex systems
- Curse of dimensionality and variable screening
- Overview of approximation methods
 - Interpolation: Kriging and radial basis functions
 - Local regression for smooth approximation
 - Robust data fitting
 - Ridge regression to avoid overfitting
 - Fitting data in feature space for irregularly distributed data
 - Variable-fidelity approximation for calibration of low-fidelity models using high-fidelity data
- Selection of model parameters
- Visualization in high-dimensional space
- Conclusions and future research problems

Interpolation of Data

- Polynomial interpolation (1D, but 2D requires nice data distribution)
- Spline interpolation (1D and 2D, any data distribution)
- Kriging interpolation: any dimension and any data distribution
- Radial basis function (RBF) interpolation: any dimension and any data distribution

History of Kriging Interpolation

- D. Krige (a South African mining engineer) developed Kriging in the 1950s to determine true ore-grades based on samples.
- Other contributors: G. Matheron (a French mathematician), L. Gandin (a Soviet meteorologist)
- Application of Kriging as approximation tool for simulation analysis: J. Sacks, W. Welch, T. Mitchell, and H. Wynn (1989)

Basics of Kriging Interpolation

- Estimating a *stationary* random process:

$$f(x) = \mu + \delta(x),$$

where μ is the mean and $\delta(x)$ is the variance.

- Kriging Interpolation:

- $\text{Var}[f(x+h)-f(x)] = \gamma(h)$

- Gaussian model: $\gamma(h) = 1 - \exp(-\|h\|^2/\sigma^2)$

- $f(x_j) = y_j$ for $1 \leq j \leq n$

- $f(x) = \sum \lambda_j(x) y_j$, $\sum \lambda_j(x) = 1$ (weighted average)

Basics of Radial Basis Function (RBF) Interpolation

- Hardy's multiquadric function for meteorological applications (1971):

$$\gamma(t) = (t^2 + a^2)^{1/2} \text{ (no statistical meaning)}$$

- Multiquadric RBF Interpolation:
 - $f(x) = \sum_{1 \leq k \leq n} c_k \gamma(\|x - x_k\|)$
 - Find c_k such that $f(x_i) = y_i$ for $1 \leq i \leq n$
- C. Micchelli (1985): RBF interpolation can be applied to any dimension and any data distribution with an appropriate choice of $\gamma(t)$.
- Other choices of $\gamma(t)$: Gaussian or $\gamma(t) = \exp(-t^2/a^2)$, linear, cubic, inverse-quadratic, and thin-plate spline

Relation Between Kriging and RBF Interpolation

Kriging interpolation can be reformulated as a constrained interpolation problem:

$$c_0 + \sum_{j=1}^n c_j \gamma(\|\mathbf{x}_i - \mathbf{x}_j\|) = y_i \quad \text{for } 1 \leq i \leq n,$$

subject to the following constraints:

$$\sum_{j=1}^n c_j = 0.$$

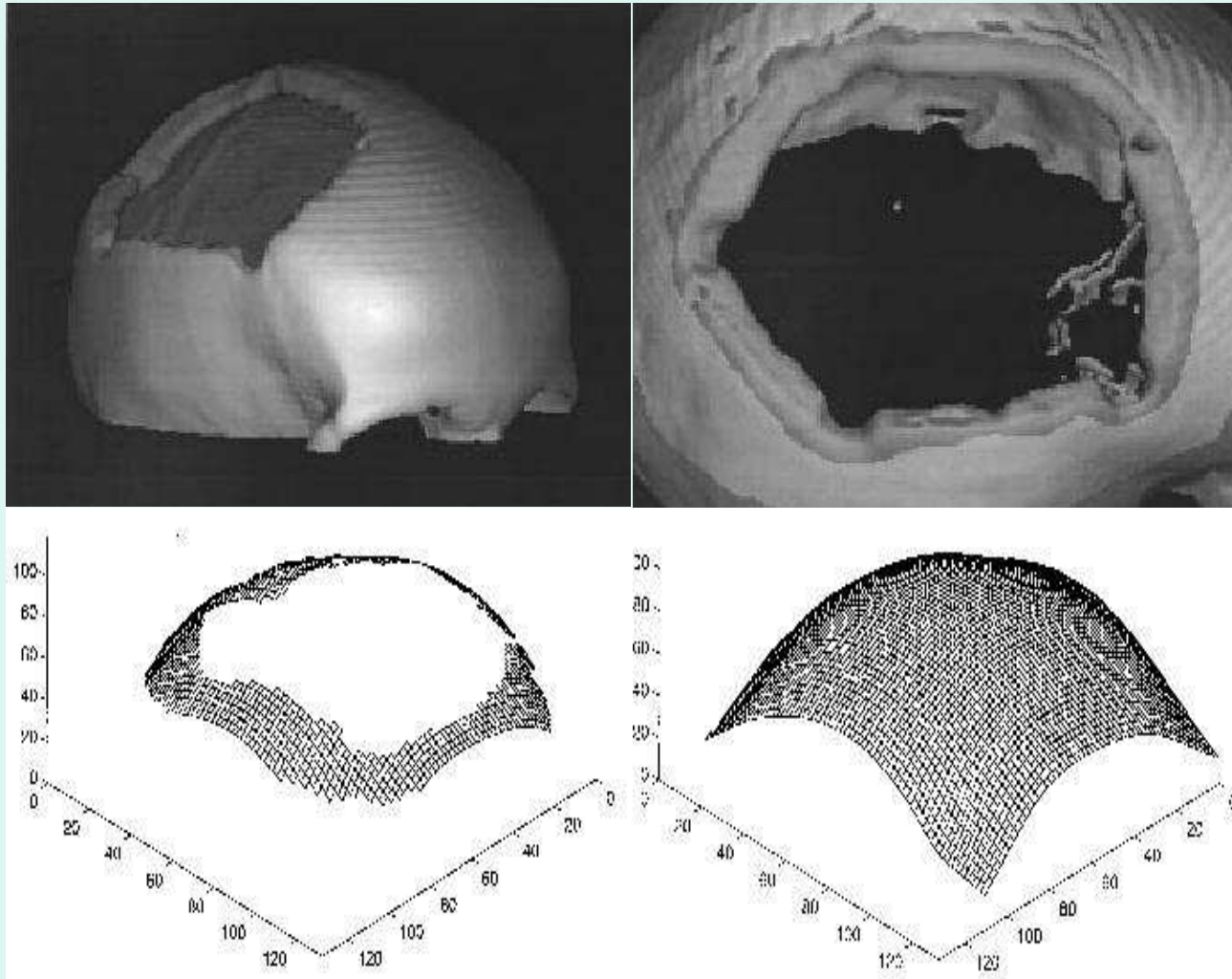
The above constrained interpolation problem was studied by numerical analysts (Powell, 1998) without any reference to Kriging interpolation.

Good News: Algorithms for RBF interpolation can be used for Kriging interpolation!

Computation of RBF Interpolation

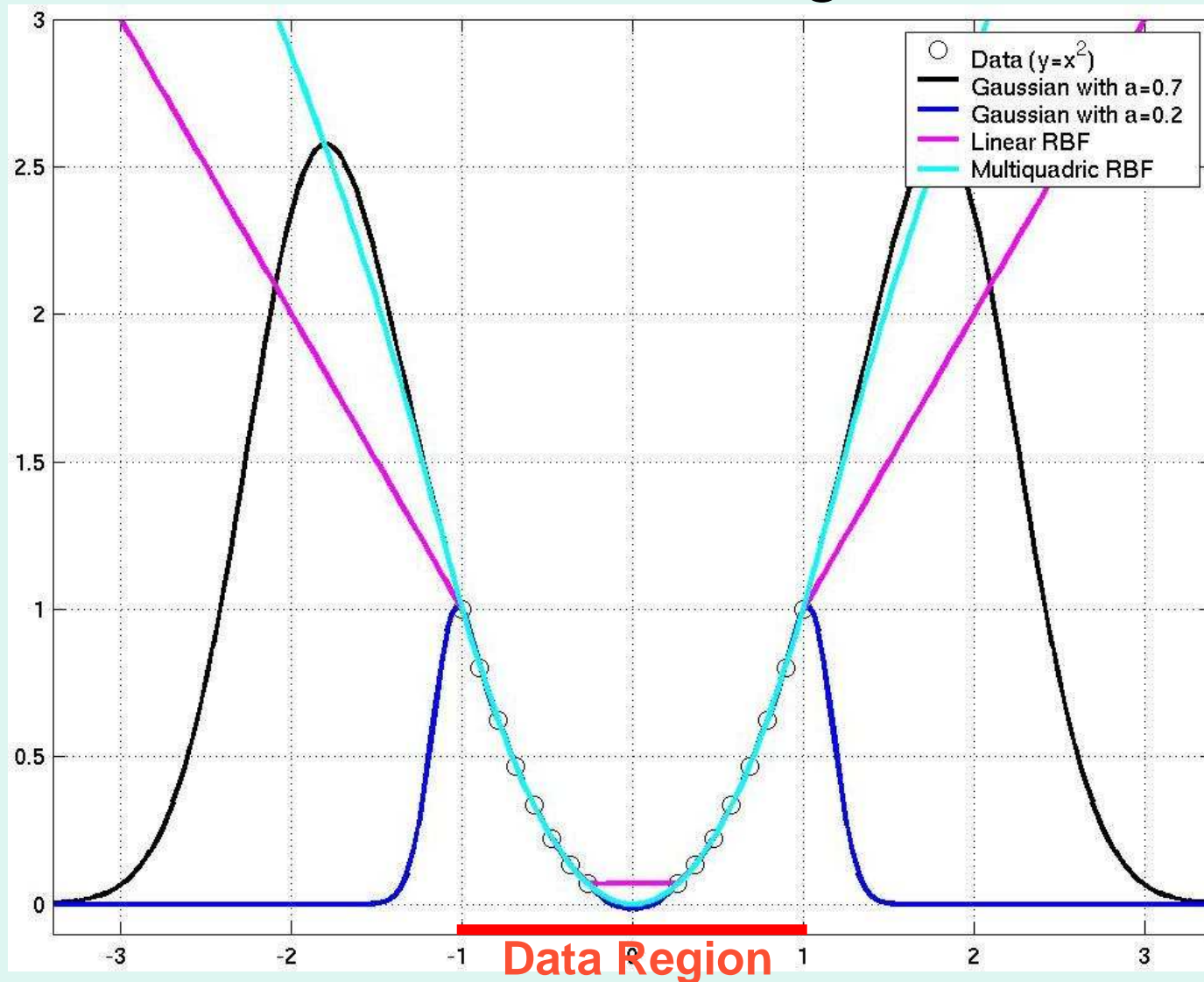
- Domain decomposition method by Beatson, Faul, Goodsell, and Powell
- Multipole method by Powell, Beatson, Light, and Newsam.
- Powell (2002) demonstrated that RBF interpolation problems with up to 45 variables and 4096 **poorly distributed** data points can be solved easily.
- Booker, Dennis, and *etal* (1998) used Kriging interpolation (with 31 variables and hundreds of **evenly distributed** data points) for helicopter rotor blade design.

RBF Interpolation for Skull Defect Repair



By J. Carr, W. Fright, and R. Beatson (1997)
IEEE Transactions on Medical Imaging

Poor Prediction of RBF Interpolation Outside Data Region



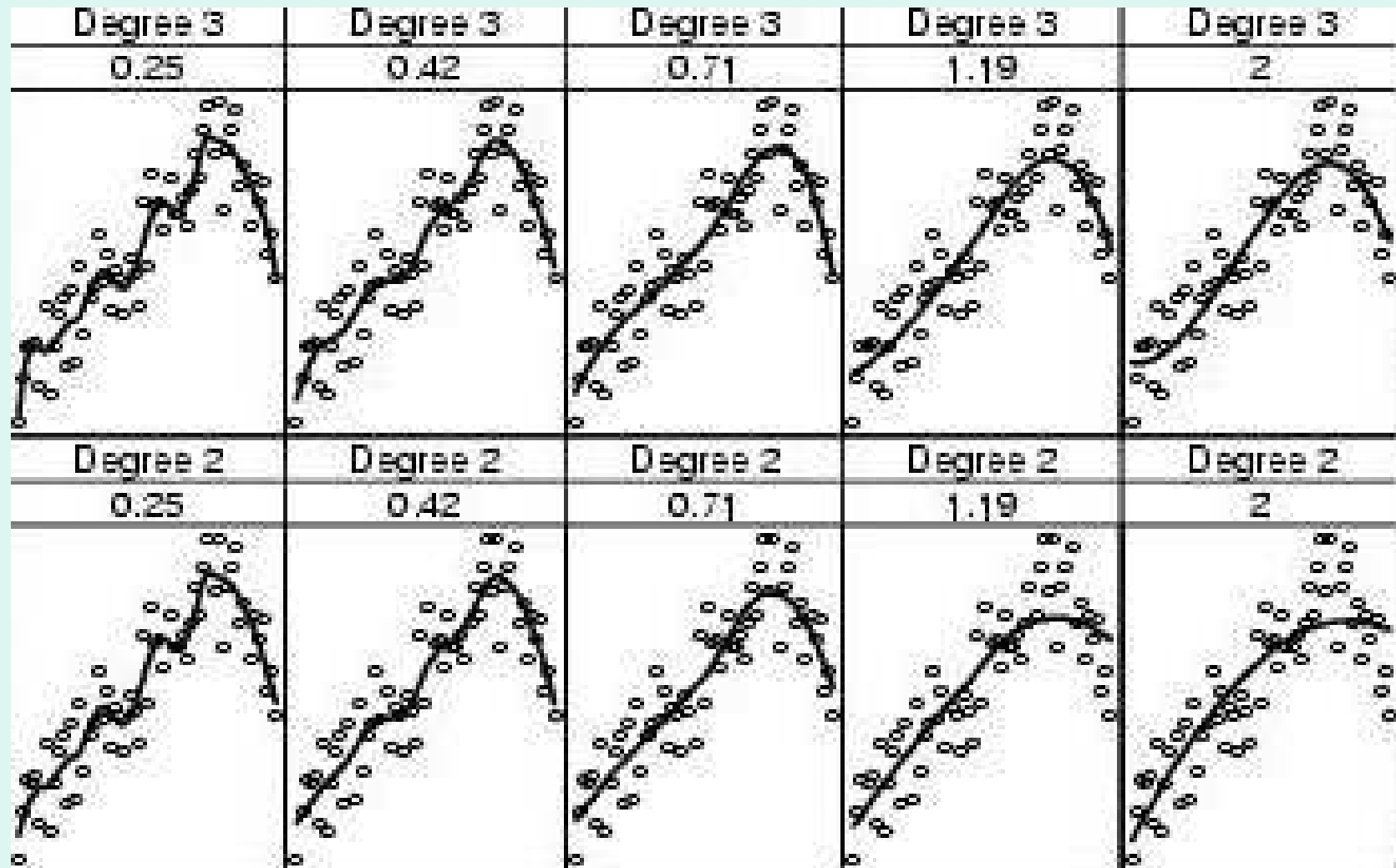
Recommended Use of RBF Interpolation in Conceptual Design

- **Historical or design data fitting**
- **Calibration of low-fidelity model**
- **Capture second order trend with a few data points**
- **Global smooth approximation**

Outline

- Needs for approximation in conceptual design of complex systems
- Curse of dimensionality and variable screening
- Overview of approximation methods
 - Interpolation: Kriging and radial basis functions
 - Local regression for smooth approximation
 - Robust data fitting
 - Ridge regression to avoid overfitting
 - Fitting data in feature space for irregularly distributed data
 - Variable-fidelity approximation for calibration of low-fidelity models using high-fidelity data
- Selection of model parameters
- Visualization in high-dimensional space
- Conclusions and future research problems

Trades Between Fitting Error and Smoothness



By Cleveland and Loader, 1996, sickness rates for ages 19 to 79 (Spencer's data, 1904)

Local Regression for Smooth Approximation

$$\min_{c_j} \sum_{i=1}^n \omega_r(\|\mathbf{x} - \mathbf{x}_i\|) \left(y_i - \sum_{j=1}^m c_j h_j(\mathbf{x}_i) \right)^2$$

Choices of the weight function include the tricube weight $\omega_r(t) = \max(0, (1 - t^3/r^3)^3)$ and Gaussian weight $\omega_r(t) = \exp(-t^2/r^2)$.

- **Locally weighted regression, local polynomial fitting:** Stone (1977) and Cleveland (1979, free software)
- **Moving least-squares:** Lancaster and Salkauskas (1981)
- **Lazy learning (in neural network community):** No explicit mathematical expression for the approximation
- **Generate accurate and smooth fitting of nonlinear responses.**
- **Require densely distributed data points.**
- **Useful for real-time design space exploration based on high-fidelity results. (Tu and Jones, 2003)**

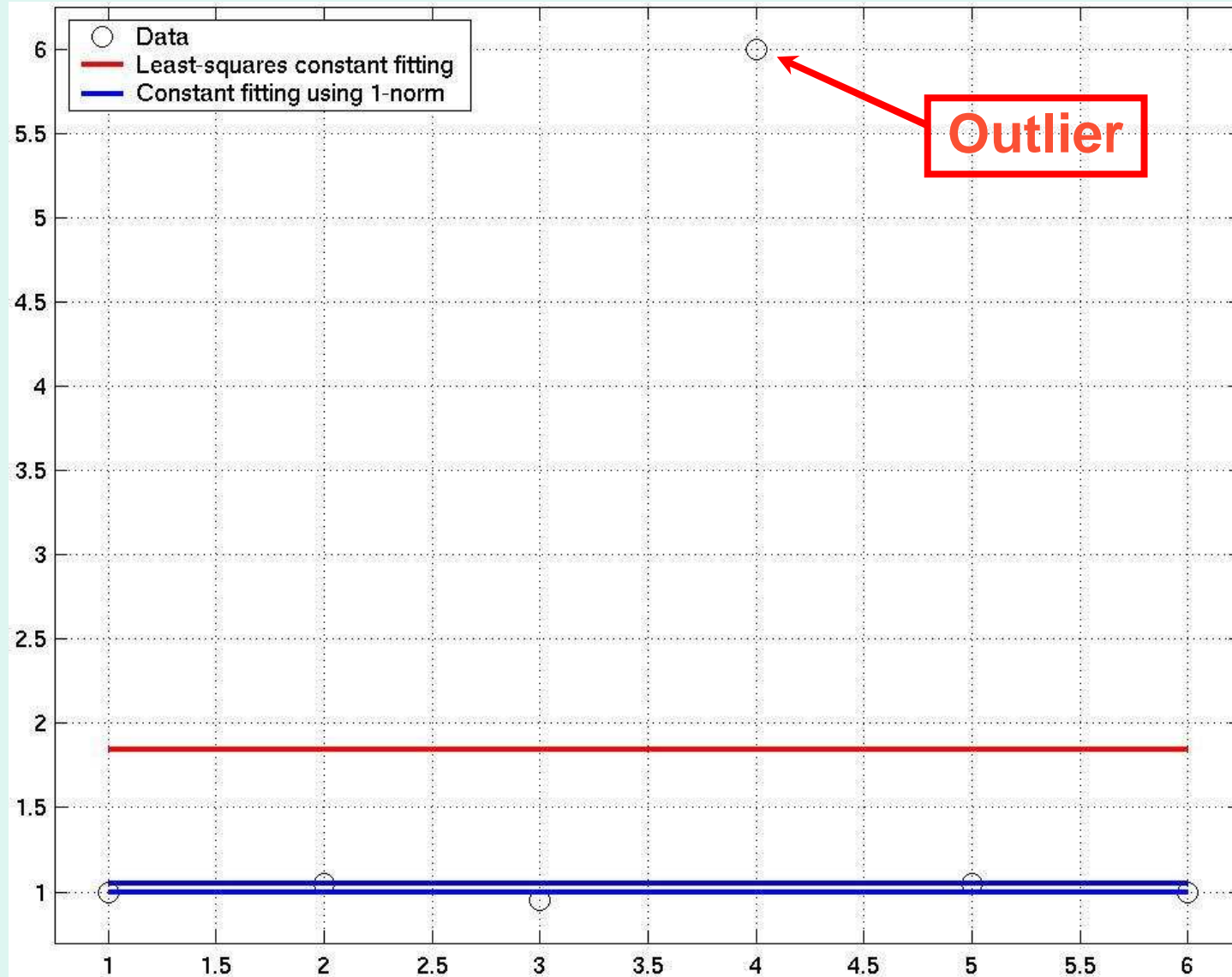
Recommended Use of Local Regression in Conceptual Design

- **Historical or design data fitting**
- **Calibration of low-fidelity model**
- **Capture second order trend with a few data points**
- **Global smooth approximation**

Outline

- Needs for approximation in conceptual design of complex systems
- Curse of dimensionality and variable screening
- Overview of approximation methods
 - Interpolation: Kriging and radial basis functions
 - Local regression for smooth approximation
 - Robust data fitting
 - Ridge regression to avoid overfitting
 - Fitting data in feature space for irregularly distributed data
 - Variable-fidelity approximation for calibration of low-fidelity models using high-fidelity data
- Selection of model parameters
- Visualization in high-dimensional space
- Conclusions and future research problems

Robust Data Fitting



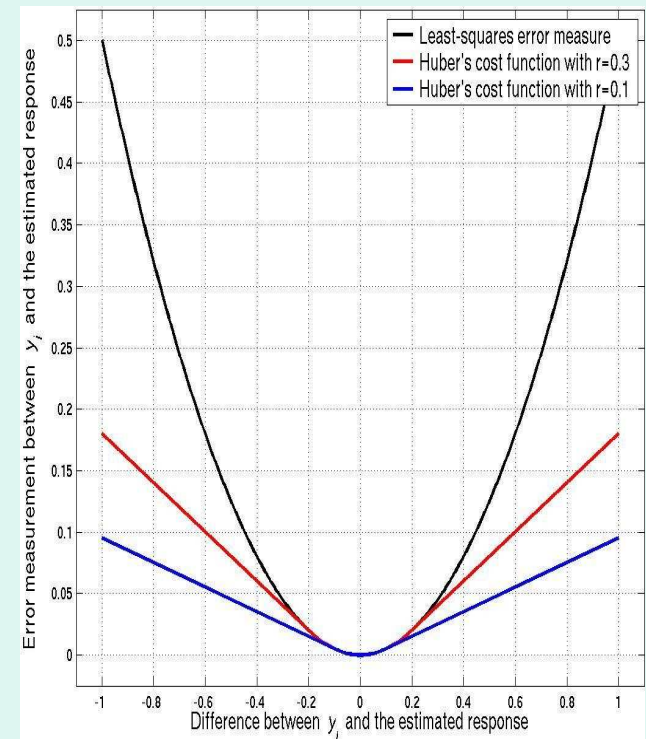
Robust Data Fitting

Robust data fitting is similar to least-squares data fitting, except the fitting error is measured by a cost function $\rho(\text{error})$ instead of squares of error:

$$\min_{c_j} \sum_{i=1}^n \rho \left(y_i - \sum_{j=1}^m c_j h_j(\mathbf{x}_i) \right)$$

Choices of ρ : 1-norm ($\rho(t)=|t|$) or Huber's cost function, *i.e.*,

$$\rho(t) = \begin{cases} \frac{1}{2}t^2, & \text{if } |t| \leq r \\ \frac{1}{2}r(2|t| - r), & \text{if } |t| > r \end{cases}$$



Control parameter r provides continuous morphing from least-squares fitting to 1-norm fitting.

Outline

- Needs for approximation in conceptual design of complex systems
- Curse of dimensionality and variable screening
- Overview of approximation methods
 - Interpolation: Kriging and radial basis functions
 - Local regression for smooth approximation
 - Robust data fitting
 - Ridge regression to avoid overfitting
 - Fitting data in feature space for irregularly distributed data
 - Variable-fidelity approximation for calibration of low-fidelity models using high-fidelity data
- Selection of model parameters
- Visualization in high-dimensional space
- Conclusions and future research problems

Ridge Regression to Avoid Overfitting ($m > n$)

- A. Hoerl and W. Kennard (1970) proposed ridge regression for automatic selection or simplification of the fitting model:

$$\min_{c_j} \sum_{i=1}^n \left(y_i - \sum_{j=1}^m c_j h_j(\mathbf{x}_i) \right)^2 + \delta \sum_{j=1}^m c_j^2$$

- R. Tibshirani (1996) demonstrated that the 1-norm penalty term does a better job:

$$\min_{c_j} \sum_{i=1}^n \left(y_i - \sum_{j=1}^m c_j h_j(\mathbf{x}_i) \right)^2 + \delta \sum_{j=1}^m |c_j|$$

The penalty term in the red box forces c_j to be zero if $h_j(\mathbf{x})$ is not useful for data fitting.

Quadratic Smoothing and Interpolation

Consider the following quadratic polynomial:

$$P_2(\mathbf{x}) = a_0 + \sum_{j=1}^k a_j x_j + \sum_{1 \leq i \leq j \leq k} b_{ij} x_i x_j.$$

If the number of data points is less than the number of coefficients, $(k+1)(k+2)/2$, in the quadratic polynomial, then we can capture some second order trend in the data by solving the following optimization problem (Powell, 2003):

$$\min_{a_i, b_{ij}} \sum_{1 \leq i \leq j \leq k} (b_{ij})^2$$

subject to the following interpolation constraints:

$$P_2(\mathbf{x}_i) = y_i \quad \text{for } 1 \leq i \leq n.$$

Similar to ridge regression, the penalty objective forces irrelevant quadratic terms to be zero. At the same time, minimizing the energy of second derivatives yields a smooth approximation.

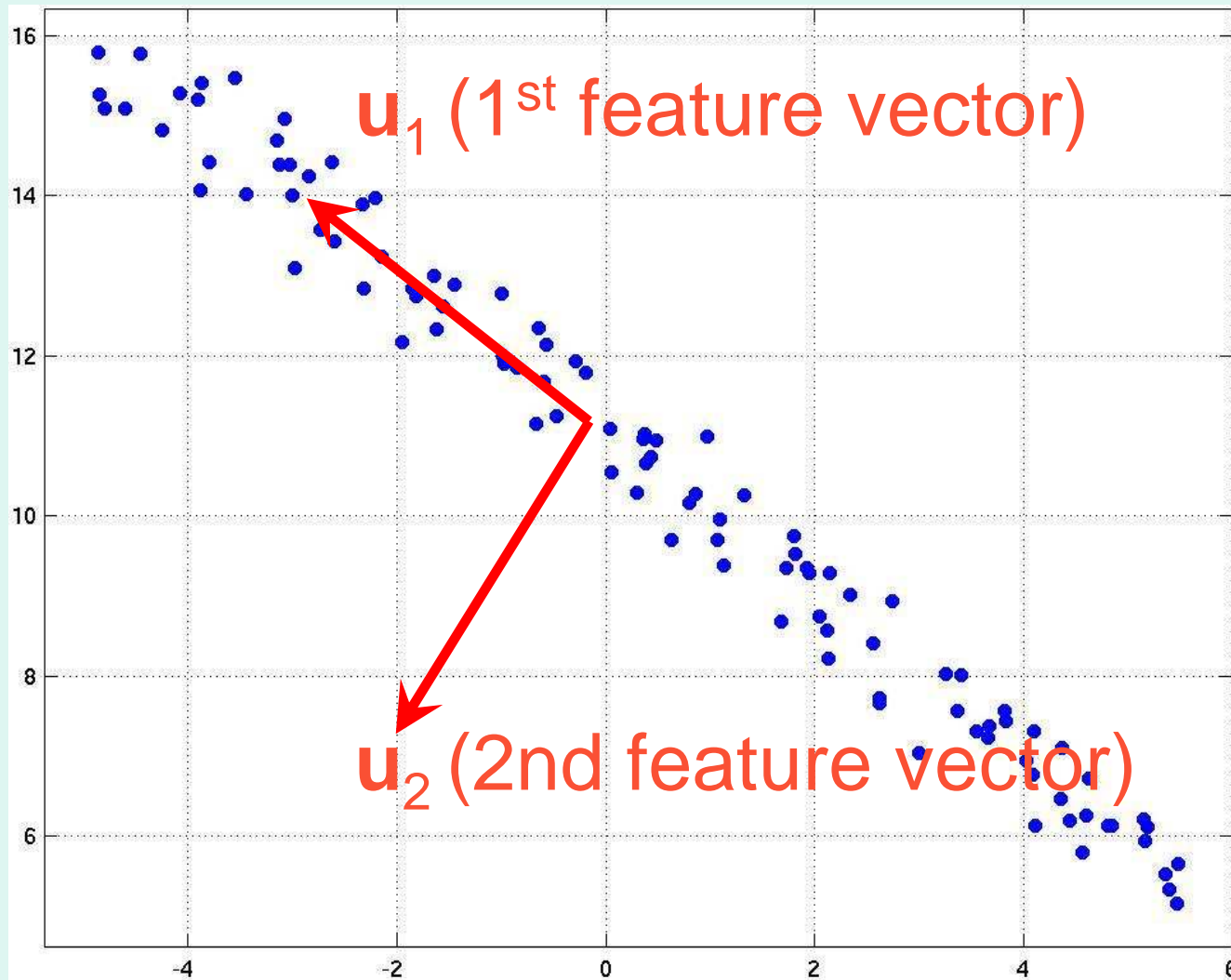
Recommended Use of Ridge Regression in Conceptual Design

- **Historical or design data fitting**
- **Calibration of low-fidelity model**
- **Capture second order trend with a few data points**
- **Global smooth approximation**

Outline

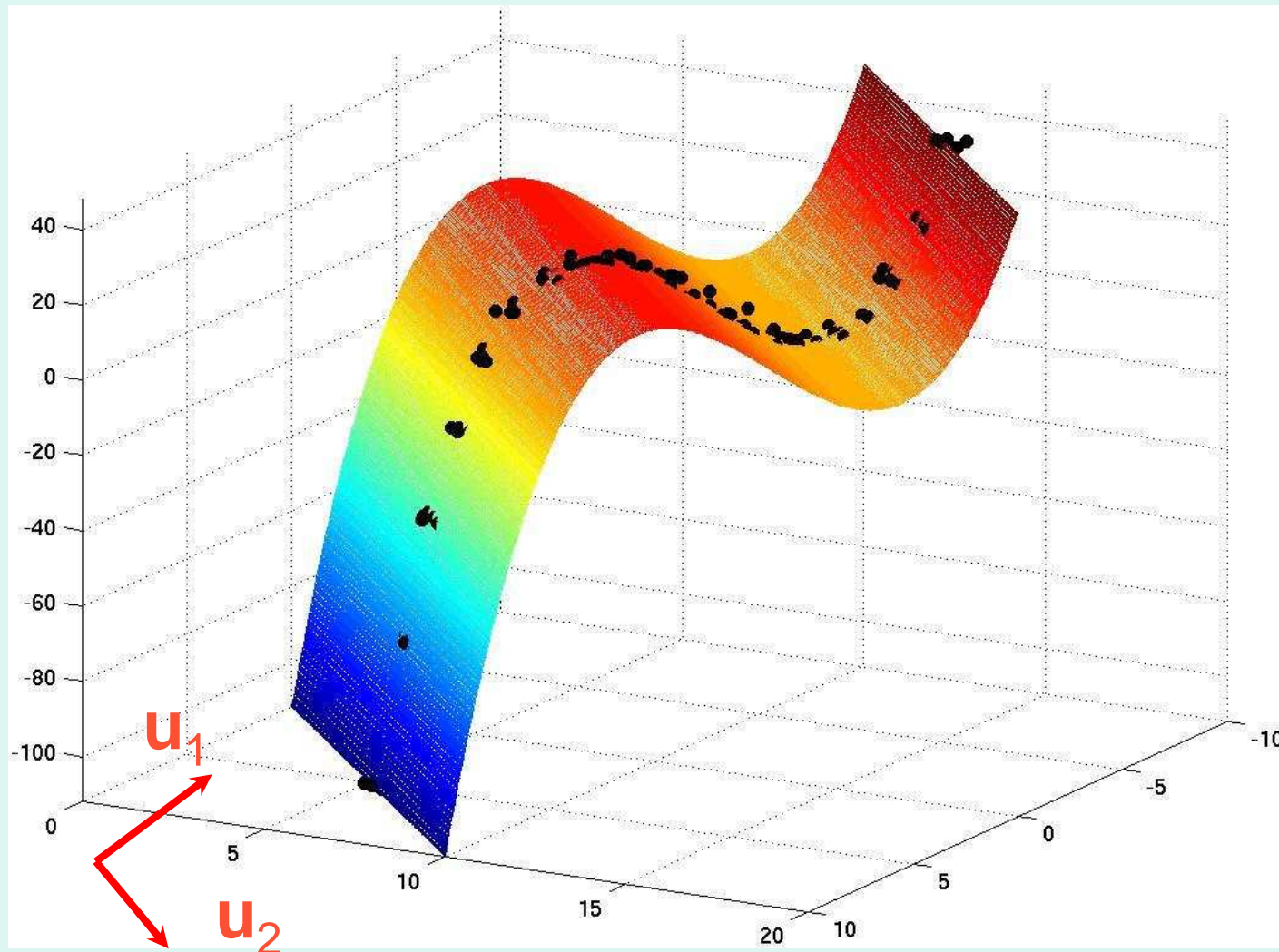
- Needs for approximation in conceptual design of complex systems
- Curse of dimensionality and variable screening
- Overview of approximation methods
 - Interpolation: Kriging and radial basis functions
 - Local regression for smooth approximation
 - Robust data fitting
 - Ridge regression to avoid overfitting
 - Fitting data in feature space for irregularly distributed data
 - Variable-fidelity approximation for calibration of low-fidelity models using high-fidelity data
- Selection of model parameters
- Visualization in high-dimensional space
- Conclusions and future research problems

Principal Component Analysis



**Irregular data distribution in the input space:
historical/measurement data, or design data**

Example of Data Fitting in Feature Space



Fitting the data along the 1st feature direction

Principal Component Regression (PCR)

- Principal Component Analysis has been widely used for dimension reduction in speech recognition (such as feature extraction of a spoken sentence) and other pattern classification problems.
- Idea of PCR was proposed by W. Massey (1965).
- PCR is a popular tool for developing calibration techniques used in near-infrared reflectance spectroscopy ...

Data Fitting in Feature Space

Input vectors can be represented as a linear combination of feature vectors (Principle Component Analysis):

$$\mathbf{x}_i - \bar{\mathbf{x}} = \sum_{j=1}^{\hat{m}} [\mathbf{u}_j^T (\mathbf{x}_i - \bar{\mathbf{x}})] \mathbf{u}_j.$$

Projection onto a reduced feature space ($m < \hat{m}$):

$$\mathcal{P}(\mathbf{x}) = \bar{\mathbf{x}} + \sum_{j=1}^m [\mathbf{u}_j^T (\mathbf{x} - \bar{\mathbf{x}})] \mathbf{u}_j.$$

Feature coordinates $\mathbf{v} = (v_1, \dots, v_m)$ for \mathbf{x} are defined by

$$v_j = \mathbf{u}_j^T (\mathbf{x} - \bar{\mathbf{x}}) \quad \text{for } 1 \leq j \leq m.$$

Construct an approximation $\hat{h}(\mathbf{v})$ in the reduced feature space:

$$\hat{h}(\mathbf{v}_i) \approx y_i \quad \text{for } 1 \leq i \leq n.$$

Recover an approximation in the original input space:

$$\hat{f}(\mathbf{x}) = \hat{h}(\mathbf{u}_1^T (\mathbf{x} - \bar{\mathbf{x}}), \dots, \mathbf{u}_m^T (\mathbf{x} - \bar{\mathbf{x}})).$$

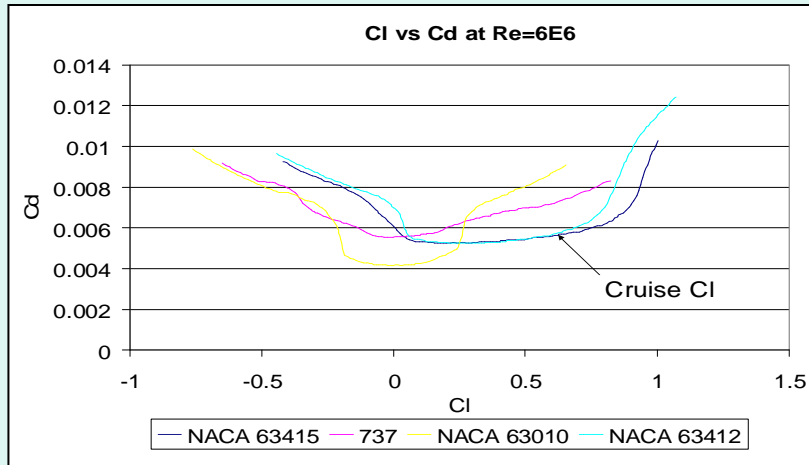
Recommended Use of PCR in Conceptual Design

- **Historical or design data fitting**
- **Calibration of low-fidelity model**
- **Capture second order trend with a few data points**
- **Global smooth approximation**

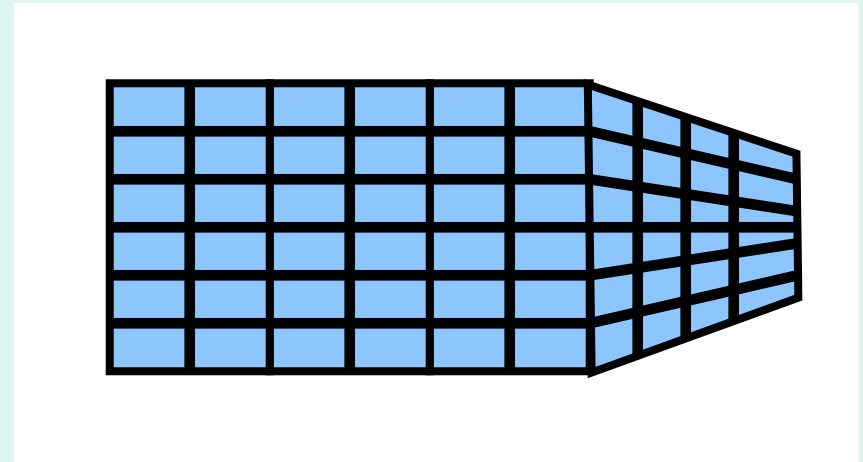
Outline

- Needs for approximation in conceptual design of complex systems
- Curse of dimensionality and variable screening
- Overview of approximation methods
 - Interpolation: Kriging and radial basis functions
 - Local regression for smooth approximation
 - Robust data fitting
 - Ridge functions or neural nets for trend prediction
 - Ridge regression to avoid overfitting
 - Fitting data in feature space for irregularly distributed data
 - Variable-fidelity approximation for calibration of low-fidelity models using high-fidelity data
- Selection of model parameters
- Visualization in high-dimensional space
- Conclusions and future research problems

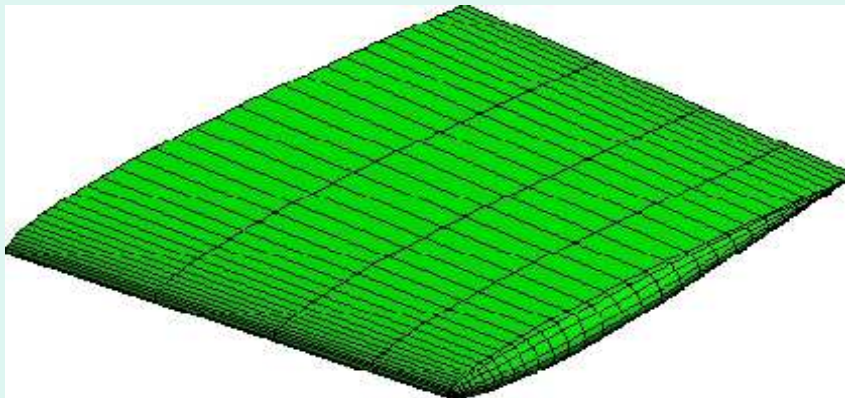
Variable-Fidelity Aerodynamics Calculation



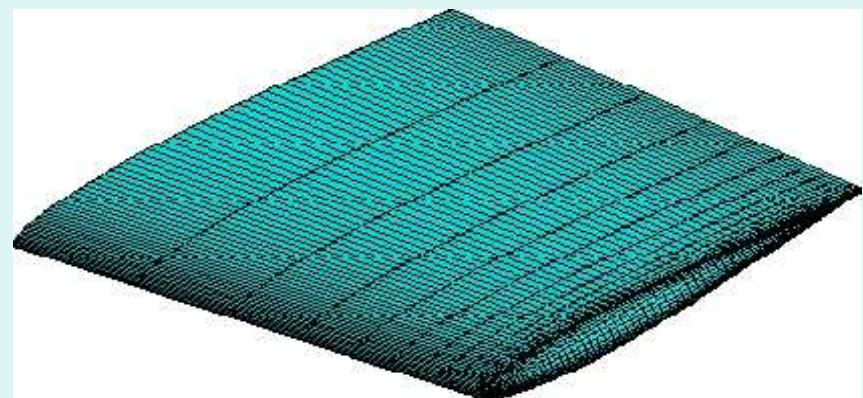
1. Drag Polar Needed



2. Panel Codes



3. Euler Codes



4. Navier-Stokes Codes

Variable Fidelity Approximation

Suppose that $y = f(\mathbf{x})$ is the true response of the systems, $y_i \approx f(\mathbf{x}_i)$ ($1 \leq i \leq n$) are high-fidelity data available, and $y = f_L(\mathbf{x})$ is a low-fidelity simulation model of $f(\mathbf{x})$.

- Additive correction of $f_L(\mathbf{x})$ to match the data:

$$f_V(\mathbf{x}) = f_L(\mathbf{x}) + \mu(\mathbf{x}),$$

where $\mu(\mathbf{x})$ is an interpolation of the data points $(\mathbf{x}_i, y_i - f_L(\mathbf{x}_i))$ ($1 \leq i \leq n$), i.e.,

$$\mu(\mathbf{x}_i) = y_i - f_L(\mathbf{x}_i) \quad \text{or} \quad f_V(\mathbf{x}_i) = y_i \quad \text{for } 1 \leq i \leq n.$$

- Multiplicative correction of $f_L(\mathbf{x})$ to match the data:

$$f_V(\mathbf{x}) = \beta(\mathbf{x})f_L(\mathbf{x}),$$

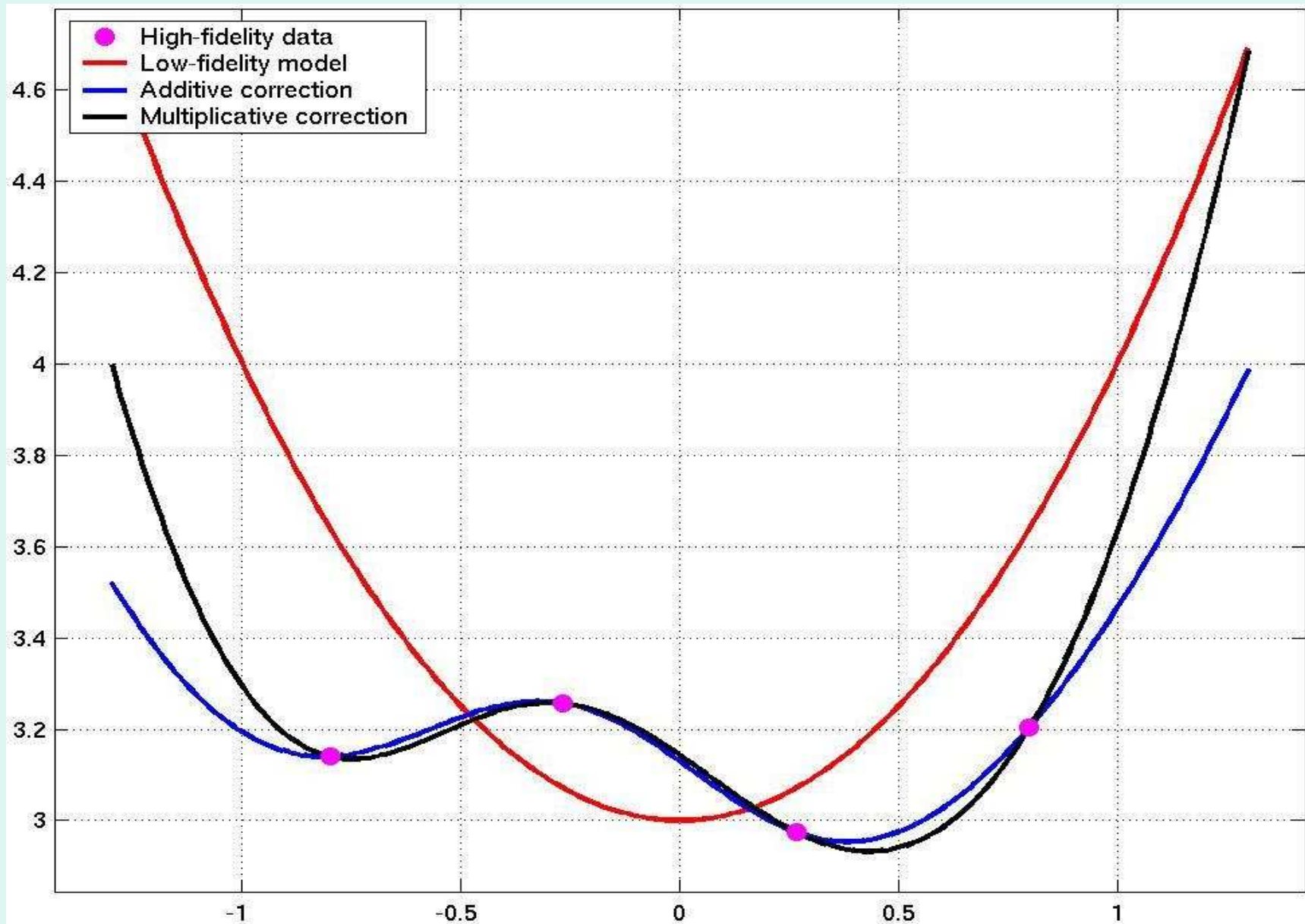
where $\beta(\mathbf{x})$ is usually an interpolation of the data points $(\mathbf{x}_i, y_i/f_L(\mathbf{x}_i))$ ($1 \leq i \leq n$), i.e.,

$$\beta(\mathbf{x}_i) = \frac{y_i}{f_L(\mathbf{x}_i)} \quad \text{or} \quad f_V(\mathbf{x}_i) = y_i \quad \text{for } 1 \leq i \leq n.$$

- Haftka's scaling function using the gradient information:

$$f_V(\mathbf{x}) = \frac{f_H(\mathbf{x}_0)}{f_L(\mathbf{x}_0)} \left[1 + (\mathbf{x} - \mathbf{x}_0)^T \left(\frac{\nabla f_H(\mathbf{x}_0)}{f_H(\mathbf{x}_0)} - \frac{\nabla f_L(\mathbf{x}_0)}{f_L(\mathbf{x}_0)} \right) \right] f_L(\mathbf{x}).$$

Illustration of Variable Fidelity Approximation



Recommended Use of Variable Fidelity Approximation in Conceptual Design

- **Historical or design data fitting**
- **Calibration of low-fidelity model**
- **Capture second order trend with a few data points**
- **Global smooth approximation**

Outline

- Needs for approximation in conceptual design of complex systems
- Curse of dimensionality and variable screening
- Overview of approximation methods
 - Interpolation: Kriging and radial basis functions
 - Local regression for smooth approximation
 - Robust data fitting
 - Ridge functions or neural nets for trend prediction
 - Ridge regression to avoid overfitting
 - Fitting data in feature space for irregularly distributed data
 - Variable-fidelity approximation for calibration of low-fidelity models using high-fidelity data
- Selection of model parameters
- Visualization in high-dimensional space
- Conclusions and future research problems

Selection of Model Parameter

- Let $\hat{f}_{j,r}(\mathbf{x})$ be the constructed approximant of the data points (\mathbf{x}_i, y_i) for $i \neq j$ with a model parameter or bandwidth r .
- Then cross-validation for least-squares fitting finds the optimal value of r that minimizes the following merit function:

$$\mathcal{CV}(r) = \sum_{j=1}^n (y_j - \hat{f}_{j,r}(\mathbf{x}_j))^2.$$

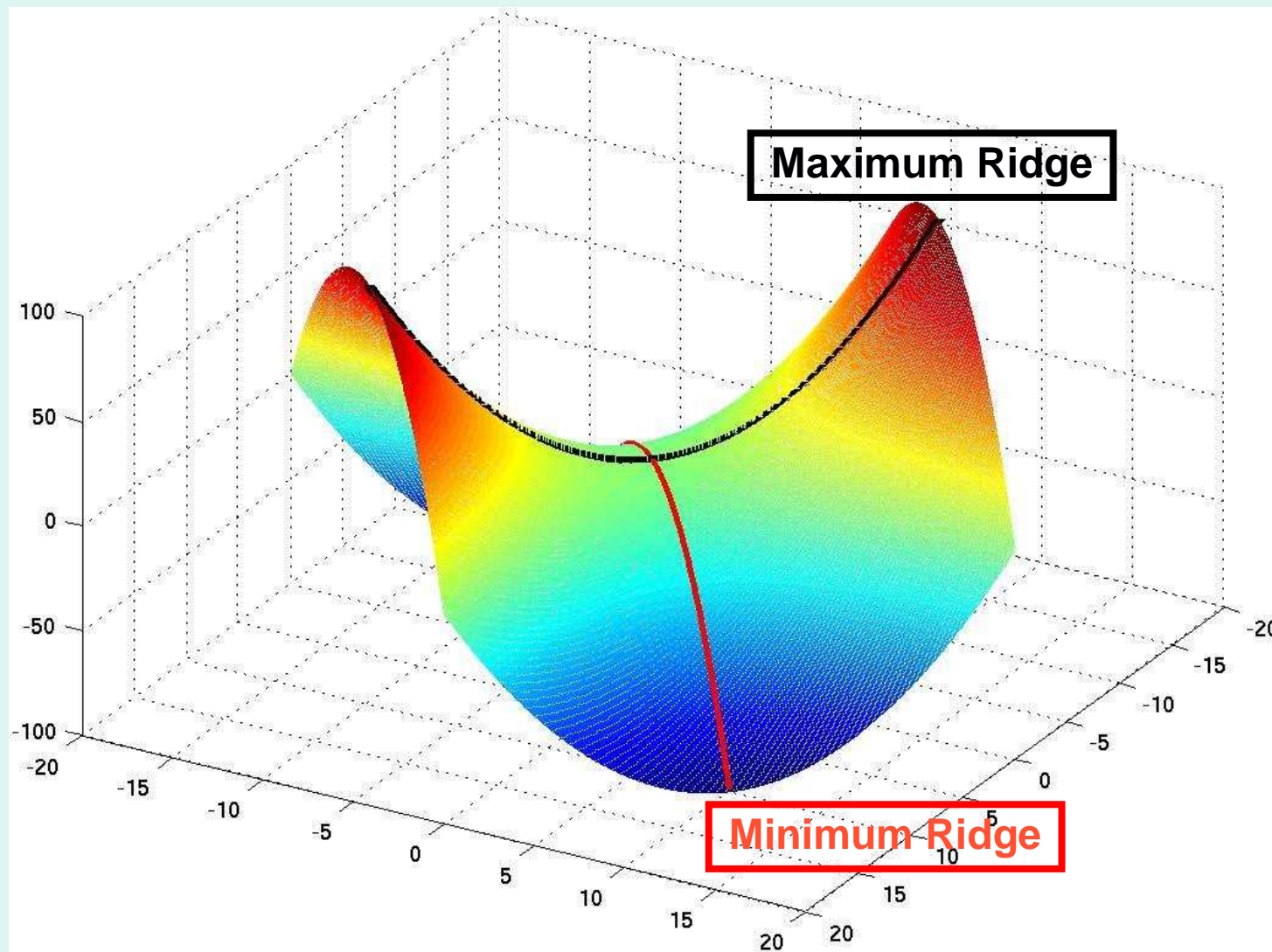
- Note that $(y_j - \hat{f}_{j,r}(\mathbf{x}_j))^2$ measures the predictive capability of $\hat{f}_{j,r}(\mathbf{x})$ at the data point (\mathbf{x}_j, y_j) that was not used for the construction of $\hat{f}_{j,r}(\mathbf{x})$.

Goal: Choose the model parameter that provides the maximum predictive capability for the given data.

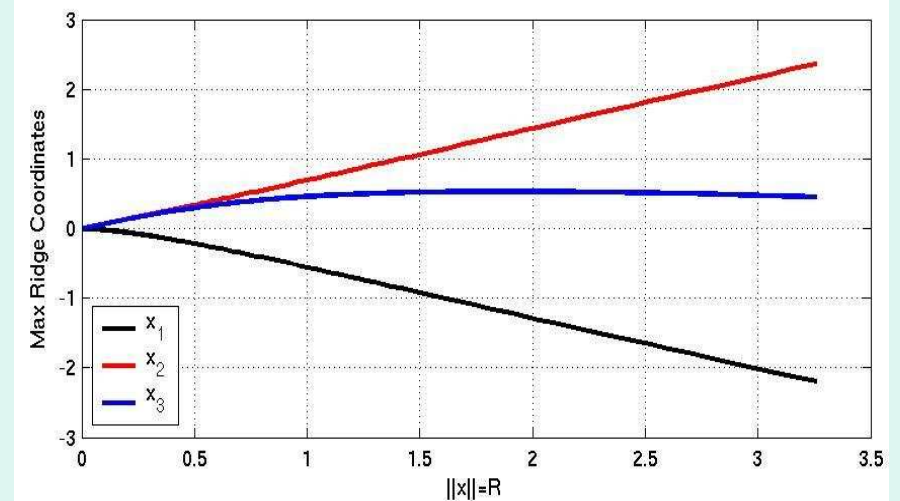
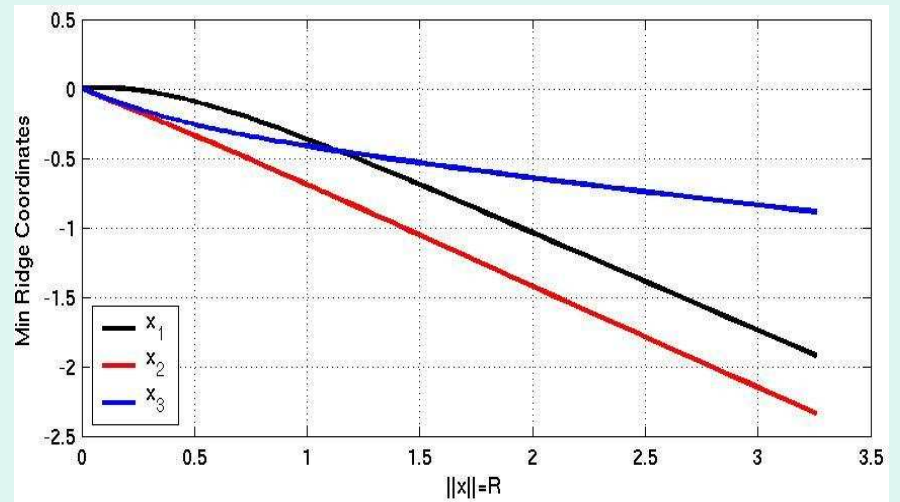
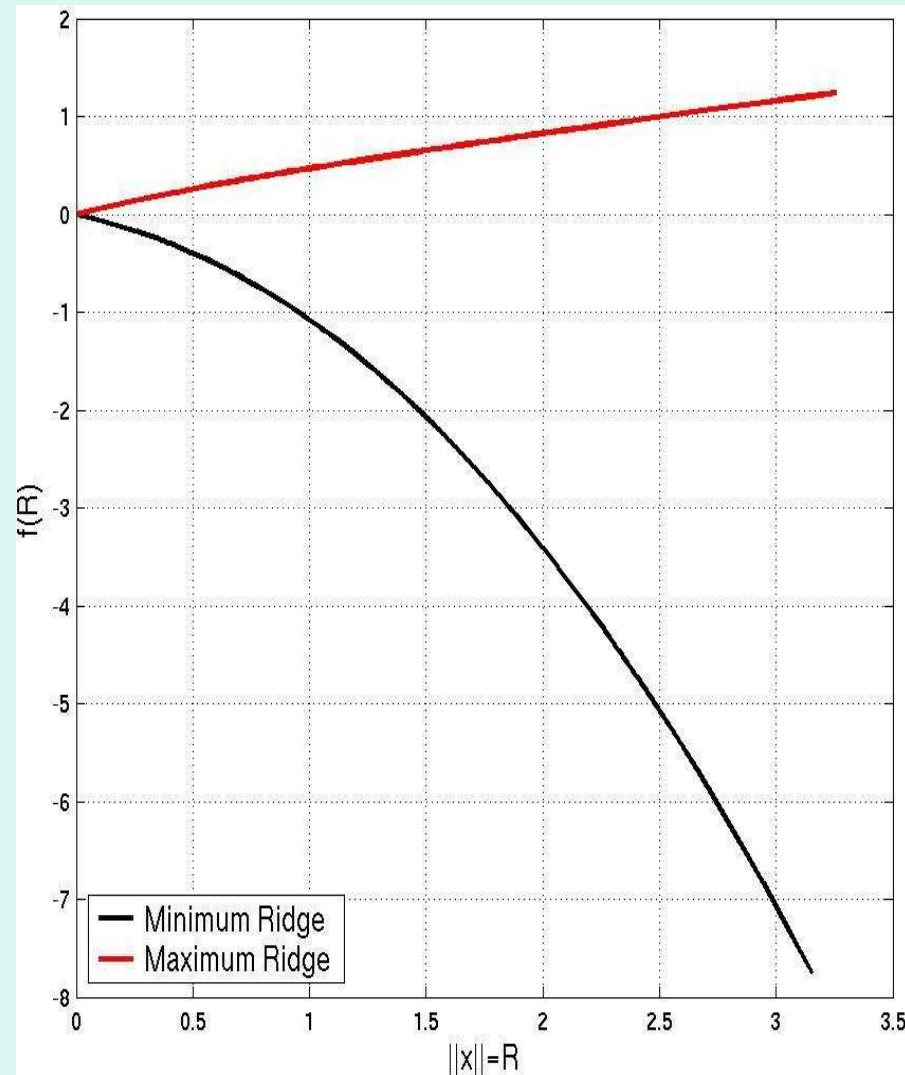
Outline

- Needs for approximation in conceptual design of complex systems
- Curse of dimensionality and variable screening
- Overview of approximation methods
 - Interpolation: Kriging and radial basis functions
 - Local regression for smooth approximation
 - Robust data fitting
 - Ridge functions or neural nets for trend prediction
 - Ridge regression to avoid overfitting
 - Fitting data in feature space for irregularly distributed data
 - Variable-fidelity approximation for calibration of low-fidelity models using high-fidelity data
- Selection of model parameters
- Visualization in high-dimensional space
- Conclusions and future research problems

Maximum and Minimum Ridges of a Quadratic Polynomial



Visualization of Quadratic Polynomial of Many Variables by Ridge Analysis (Hoerl)



Concluding Remarks

- **Historical or design data fitting**
 - RBF fitting
 - Data fitting in feature space
- **Capture second order trend with a few data points**
 - Ridge regression
 - Powell's smooth quadratic interpolation
- **Calibration of low-fidelity model**
 - Multiplicative correction
 - Additive correction
- **Global smooth approximation**
 - Local Regression
- **Visualization: Maximum and minimum ridges**

Future Research Problems

- Multiplicative or additive correction that preserves some physical characteristics of the high-fidelity data (such as the scaling grows linearly or quadratically outside the data region)
- Interpolation that allows inhomogeneous directional trend prediction
- Given a quadratic function $q(x)$ and an integer m , find vectors u_k and univariate quadratic polynomial $p_k(t)$ such that $\sum_{1 \leq k \leq m} p_k(u_k \bullet x)$ approximates $q(x)$ the best on a compact convex set.
- Mathematical theory on the smoothness of local polynomial fitting

Acknowledgments

We would like to thank Andy Hahn, Tom Ozoroski, and Arnie McCullers for several stimulating discussions that help us understand some critical needs of approximation methods in systems analysis.

See further Info at <http://mdob.larc.nasa.gov>